MATH 568 ANALYTIC NUMBER THEORY I.
CHAPTER 0

0.1. **Introduction.** Number theory in its most basic form is the study of the set of *integers*

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$$

and its important subset

$$\mathbb{N} = \{1, 2, 3, \dots\}.$$

Analytic number theory is the use of analytic techniques to understand the properties of these sets. For example Waring's problem (1770) is that of trying to find the smallest s, say $g(k)$, such that every positive integer is the sum of at most s k-the powers. One way of doing this might be to look at

$$f(z)^s = \left(\sum_{n=0}^{\infty} z^{n^k} \right)^s.$$

This converges absolutely when $|z| < 1$. Expanding it and rearranging the terms gives

$$\sum_{n=0}^{\infty} R_s(n) z^n$$

where $R_s(n)$ is the number of solutions of

$$n_1^k + \cdots + n_s^k$$

in non-negative integers $n_1, \dots, n_s$. Thus if we could show that for some $s = s_0$ we have $R_s(n) > 0$ for every n, then it would follow that

$$g(k) \leq s_0.$$

One way of doing this would be to use Cauchy's integral formula

$$R_s(n) = \frac{1}{2\pi i}\int_{\mathcal{C}} f(z)^s z^{-n-1} dz$$

and to try and evaluate the integral, at least approximately. This is the genesis of the Hardy-Littlewood-Ramanujan method and has lead to a great deal of progress on additive problems.

Another kind of question where analysis can play a rôle is in distributional questions. For example, how are the prime numbers distributed? In particular, if $\pi(x)$ is the number of primes not exceeding x, how does $\pi(x)$ behave as x grows. This is connected with the Riemann Hypothesis (RH) concerning the zeros of the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Here s is a complex number and the above converges absolutely when $\Re s > 1$. However a meaning can be assigned to the function whenever $s \neq 1$. We know that it has zeros at the negative even integers, and the RH says that all its other zeros have real part $\frac{1}{2}$. This is the most important unsolved problem in mathematics and is the first Millennium Problem. We plan to discuss some of the properties of $\zeta(s)$ later in the course.

I presume that everyone is familiar with Euclid's proof that there are infinitely many primes. That is, argue by contradiction and so suppose there are only a finite number of them, say $p_1, p_2, \ldots, p_n$ and consider $m = p_1p_2\cdots p_n + 1$. This is a prime or a product of primes, so is divisible by some prime p, say. But p also divides $p_1p_2\ldots p_n$, so $p|m - p_1p_2\ldots p_n = 1$, which is impossible.

Well here is a proof of the infinitude of primes which is essentially due to Euler and is analytic in nature. To begin first consider the important sum

$$S(x)=\sum_{n=1}^{x}\frac{1}{n}$$

where x is a large real number. Of course the sum behaves a bit like the integral so is a bit like $\log x$. In fact there is something more precise which one can say, which was discovered by Euler. Throughout for any real number y we use $\lfloor y\rfloor$ to denote the largest integer not exceeding y. We have

$$\begin{aligned}S(x)&=\sum_{n\le x}\left(\frac{1}{x}+\int_n^x\frac{dt}{t^2}\right)=\frac{\lfloor x\rfloor}{x}+\int_1^x\frac{\lfloor t\rfloor}{t^2}dt\\&=\int_1^x\frac{1}{t}dt+1-\int_1^x\frac{t-\lfloor t\rfloor}{t^2}dt-\frac{x-\lfloor x\rfloor}{x}\\&=\log x+C_0+\int_1^x\frac{t-\lfloor t\rfloor}{t^2}dt-\frac{x-\lfloor x\rfloor}{x}\end{aligned}$$

which gives

$$\sum_{n\le x}\frac{1}{n}=\log x+C_0+O(1/x). \tag{0.1}$$

Here $C_0=0.577\ldots$ is Euler's constant, which he computed to 19 decimal places (by hand of course). Actually that is not so hard and we might say something about it later.

By the way, it would be good here to say something about notation. Typically most latin letters will be integers or natural numbers, but t, x, y may well be real numbers, according to context, and z, and in Dirichlet series s, will be complex numbers.

Given functions f and g defined on some domain $\mathcal{X}$ with $g(x)\ge 0$ for all $x\in\mathcal{X}$ we write

$$f(x)=O(g(x)) \tag{0.2}$$

to mean that there is some constant C such that

$$|f(x)| \leq Cg(x)$$

for every $x \in \mathcal{X}$. We also use

$$f(x) = o(g(x)$$

to mean that if there is some limiting operation, such as $x \to \infty$, then

$$\frac{f(x)}{g(x)} \to 0$$

and

$$f(x) \sim g(x)$$

to mean

$$\frac{f(x)}{g(x)} \to 1.$$

The symbols O and o were invented by Bachmann, Landau's doctoral supervisor about 120 years ago. The O-symbol can be a bit clumsy for complicated expressions and we will often instead use the Vinogradov symbols, which I. M. Vinogradov introduced about 90 years ago. Thus we will use

$$f \ll g \tag{0.3}$$

to mean (0.2). This has the advantage that we can write strings of inequalities in the form

$$f_1 \ll f_2 \ll f_3 \ll \dots.$$

Also if f is also non-negative we may use

$$g \gg f$$

to mean (0.3).

Return to $S(x)$. Less precise than Euler's result is the observation that

$$S(x) \geq \sum_{n \leq x} \int_n^{n+1} \frac{dt}{t} \geq \int_1^x \frac{dt}{t} = \log x.$$

Now consider

$$P(x) = \prod_{p \le x} (1 - 1/p)^{-1}$$

where the product is over the primes not exceeding x. Then

$$P(x) = \prod_{p \le x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots \right) \ge \sum_{n \le x} \frac{1}{n} \ge \log x.$$

Note that when one multiplies out the left hand side every fraction $\frac{1}{n}$ with $n \le x$ occurs. Since $\log x \to \infty$ as $x \to \infty$, there have to be infinitely many primes. Actually one can get something a bit more precise. Take logs on both sides. Thus

$$-\sum_{p \le x} \log(1 - 1/p) \ge \log \log x.$$

Moreover the expression on the left is

$$\sum_{p \le x} \sum_{k=1}^{\infty} \frac{1}{kp^k}.$$

Here the terms with $k \ge 2$ contribute at most

$$\sum_{p \le x} \frac{1}{2} \sum_{k=2}^{\infty} \frac{1}{p^k} \le \frac{1}{2} \sum_{n=2}^{\infty} \frac{1}{n(n-1)} = \frac{1}{2}.$$

Hence we have just proved that

$$\sum_{p \le x} \frac{1}{p} \ge \log \log x - \frac{1}{2}.$$

This is quite close to the truth, and we will show in a while that there is a constant C_1 such that

$$\sum_{p \le x} \frac{1}{p} = \log \log x + C_1 + o(1).$$

Since

$$\int_2^x \frac{dt}{t \log t} = \log \log x - \log \log 2$$

it suggests that about $1/\log n$ of the numbers near n are prime, or in other words the “probability” that n is prime is $1/\log n$. Hence one might guess that $\pi(x)$ is indeed about

$$\int_2^x \frac{dt}{\log t}$$

and the following table indicates that this is indeed true for x out to about $10^2 2$.

Show pi.jpg here. Euler’s result on primes is often quoted as follows.

Theorem 0.1 (Euler)**.** *The sum*

$$\sum_p \frac{1}{p}$$

diverges.

Since we are not sure of the number theory background of everyone in the class we will start by discussing some useful topics from elementary number theory.

0.2. **Arithmetical functions.** The set $\mathcal{A}$ of arithmetical functions is defined by

$$\mathcal{A} = \{f : \mathbb{N} \to \mathbb{C}\}.$$

Of course the range of any particular function might well be a subset of $\mathbb{C}$. The function $R_s(n)$ defined earlier, when restricted to the positive integers is such a function. There are quite a number of important arithmetical functions. Some examples are

The divisor function. The number of positive divisors of n.

$$d(n) = \sum_{m|n} 1.$$

Euler's function. The number $\phi(n)$ of integers m with $1 \leq m \leq n$ and $(m, n) = 1$. This is important because it counts the number of units in $\mathbb{Z}/n\mathbb{Z}$.

Euler's function satisfies an interesting relationship.

Theorem 0.2. *We have*

$$\sum_{m|n} \phi(m) = n.$$

One way of seeing this is as follows. Consider the n fractions

$$\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}.$$

Then factor out any common factors between denominators and numerators. Then one will obtain each fraction of the form

$$\frac{l}{m}$$

with $m|n$, $1 \leq l \leq m$ and $(l, m) = 1$. The number of such fractions is

$$\sum_{m|n} \phi(m).$$

The Möbius function. This is a more peculiar function. It is defined by

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{if there is a prime } p \text{ such that } p^2|n. \end{cases}$$

It is also convenient to introduce three very boring functions.

The unit.

$$e(n) = \begin{cases} 1 & (n = 1), \\ 0 & (n > 1). \end{cases}$$

The one.

$$\mathbf{1}(n) = 1 \text{ for every } n.$$

The identity.

$$N(n) = n.$$

Two other functions which have interesting structures but which we will say less about at this stage are

The primitive character modulo 4. We define

$$\chi_1(n) = \begin{cases} (-1)^{\frac{n-1}{2}} & 2 \nmid n, \\ 0 & 2|n. \end{cases}$$

Sums of two squares. We define $r(n)$ to be the number of ways of writing n as the sum of two squares of integers.

For example, $1 = 0^2 + (\pm 1)^2 = (\pm 1)^2 + 1^2$, so $r(1) = 4$, $r(3) = r(6) = r(7) = 0$, $r(9) = 4$, $65 = (\pm 1)^2 + (\pm 8)^2 = (\pm 4)^2 + (\pm 7)^2$ so $r(65) = 16$.

d, ϕ, e, $\mathbf{1}$, N, χ_1 have an interesting property. That is they are multiplicative.

Definition An arithmetical function f which is not identically 0 is **multiplicative** when it satisfies

$$f(mn) = f(m)f(n)$$

whenever $(m, n) = 1$. Let $\mathcal{M}$ denote the set of multiplicative functions.

The function $r(n)$ is not multiplicative, since $r(65) = 16$ but $r(5) = r(13) = 8$. Indeed the fact that $r(1) \neq 1$ would contradict the next theorem. However it is true that $r(n)/4$ is multiplicative, but this is a little trickier to prove.

Theorem 0.3. *Suppose that $f \in \mathcal{M}$. Then $f(1) = 1$.*

Proof. Since f is not identically 0 there is an n such that $f(n) \neq 0$. Hence $f(n) = f(n \times 1) = f(n)f(1)$, and the conclusion follows. □

It is pretty obvious that e, $\mathbf{1}$ and N are in $\mathcal{M}$, and it is actually quite easy to show

Theorem 0.4. *We have $\mu \in \mathcal{M}$.*

Proof. Suppose that $(m, n) = 1$. If $p^2|mn$, then $p^2|m$ or $p^2|n$, so $\mu(mn) = 0 = \mu(m)\mu(n)$. If

$$m = p_1 \dots p_k, \quad n = p'_1 \dots p'_l$$

with the p_i, p'_j distinct, then

$$\mu(mn) = (-1)^{k+l} = (-1)^k(-1)^l = \mu(m)\mu(n).$$

□

The following is very useful.

Theorem 0.5. *Suppose the $f \in \mathcal{M}$, $g \in \mathcal{M}$ and h is defined for each n by*

$$h(n) = \sum_{m|n} f(m)g(n/m).$$

Then $h \in \mathcal{M}$.

Proof. Suppose $(n_1, n_2) = 1$. Then a typical divisor m of $n_1 n_2$ is uniquely of the form $m_1 m_2$ with $m_1|n_1$ and $m_2|n_2$. Hence

$$\begin{aligned} h(n_1 n_2) &= \sum_{m_1|n_1} \sum_{m_2|n_2} f(m_1 m_2) g(n_1 n_2/(m_1 m_2)) \\ &= \sum_{m_1|n_1} f(m_1) g(n_1/m_1) \sum_{m_2|n_2} f(m_2) g(n_2/m_2). \end{aligned}$$

□

This enables is to establish an interesting property of the Möbius function.

Theorem 0.6. *We have*

$$\sum_{m|n} \mu(m) = e(n).$$

Proof. By the previous theorem the sum here is

$$\sum_{m|n} \mu(m)\mathbf{1}(n/m)$$

is in $\mathcal{M}$. Moreover if $k \geq 1$, then

$$\sum_{m|p^k} \mu(m) = \mu(1) + \mu(p) = 1 - 1 = 0$$

□

This suggests a general way of defining new functions.

Definition. Given two arithmetical functions f and g we define the **Dirichlet convolution** $f * g$ to be the function defined by

$$(f * g)(n) = \sum_{m|n} f(m)g(n/m).$$

Note that this operation is commutative - simply replace m by n/m.

It is also quite easy to see that

$$(f * g) * h = f * (g * h).$$

Write the left hand side as

$$\sum_{m|n} \left(\sum_{l|m} f(l)g(m/l) \right) h(n/m)$$

and interchange the order of summation and replace m by kl.

Dirichlet convolution has some interesting properties

1. $f * e = e * f = f$ for any $f \in \mathcal{A}$, so e is really acting as a unit.

2. $\mu * \mathbf{1} = \mathbf{1} * \mu = e$, so μ is the inverse of $\mathbf{1}$, and *vice versa.*

3. $d = \mathbf{1} * \mathbf{1}$, so $d \in \mathcal{M}$. Hence

4. $d(p^k) = k + 1$ and $d(p_1^{k_1} \dots p_r^{k_r}) = (k_1 + 1) \dots (k_r + 1)$.

Theorem 0.7 (Möbius inversion I)**.** *Suppose that $f \in \mathcal{A}$ and $g = f * \mathbf{1}$. Then $f = g * \mu$.*

Proof. We have

$$g * \mu = (f * \mathbf{1}) * \mu = f * (\mathbf{1} * \mu) = f * e = f.$$

□

Theorem 0.8 (Möbius inversion II)**.** *Suppose that $g \in \mathcal{A}$ and $f = g * \mu$, then $g = f * \mathbf{1}$.*

The proof is similar.

Theorem 0.9. *We have $\phi = \mu * N$ and $\phi \in \mathcal{M}$. Moreover*

$$\phi(n) = n \sum_{m|n} \frac{\mu(m)}{m} = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

Proof. We saw in Theorem 0.2 that $\phi * \mathbf{1} = N$. Hence by the previous theorem we have $\phi = N * \mu = \mu * N$. Therefore, by Theorem 0.5, $\phi \in \mathcal{M}$. Moreover $\phi(p^k) = p^k - p^{k-1}$ and we are done. □

Theorem 0.10. *Let $\mathcal{D} = \{f \in \mathcal{A} : f(1) \neq 0\}$. Then $\langle \mathcal{D}, * \rangle$ is an abelian group.*

Proof. Of course e is the unit, and closure is obvious. We already checked commutativity and associativity. It remains, given $f \in \mathcal{D}$, to construct an inverse. Define g iteratively by $g(1) = 1/f(1)$, $g(n) = -\sum_{\substack{m|n \\ m>1}} f(m)g(n/m)/f(1)$ and it is clear that $f * g = e$. □

0.3. **Averages of arithmetical functions.** One of the most powerful techniques we have is to take an average. One of the more famous theorems of this kind is

Theorem 0.11 (Dirichlet)**.** *Suppose that $X \in \mathbb{R}$ and $X \geq 2$. Then*

$$\sum_{n\le X} d(n) = X \log X + (2C_0 - 1)X + O(X^{1/2}).$$

Proof. We follow Dirichlet's proof method, which has become known as the *method of the parabola.* The divisor function $d(n)$ can be thought of as the number of ordered pairs of positive integers m, l such that $ml = n$. Thus when we sum over $n \le X$ we are just counting the number of ordered pairs m, l such that $ml \le X$. In other words we are counting the number of *lattice points* m, l under the rectangular hyperbola

$$xy = X.$$

We could just crudely count, given $m \le X$, the number of choices for l, namely

$$\left\lfloor \frac{X}{m} \right\rfloor$$

and obtain

$$\sum_{m\le X} \frac{X}{m} + O(X)$$

but this gives a much weaker error term.

Dirichlet's idea is to divide the region under the hyperbola into two parts. That with

$$m \le \sqrt{X},\, l \le \frac{X}{m}$$

and that with

$$l \le \sqrt{X},\, m \le \frac{X}{l}.$$

Clearly each region has the same number of lattice points. However the points m, l with $m \le \sqrt{X}$ and $l \le \sqrt{X}$ are counted in both regions. Thus we obtain

$$\begin{aligned}\sum_{n\le X} &= 2\sum_{m\le\sqrt{X}} \left\lfloor \frac{X}{m} \right\rfloor - \lfloor\sqrt{X}\rfloor^2 \\ &= 2\sum_{m\le\sqrt{X}} \frac{X}{m} - X + O(X^{1/2}) \\ &= 2X\big(\log(\sqrt{X}) + C_0\big) - X + O(X^{1/2}).\end{aligned}$$

where in the last line we used Euler's estimate (0.1). □

One can also compute an average for Euler's function

Theorem 0.12. *Suppose that $x \in \mathbb{R}$ and $x \ge 2$. Then*

$$\sum_{n\le x} \phi(n) = \frac{x^2}{2}\sum_{m=1}^{\infty} \frac{\mu(m)}{m^2} + O(x\log x).$$

We remark that the infinite series here is "well known" to be $\frac{6}{\pi^2}$.

Proof. We leave the proof largely to the class as homework. Hint: Use $\phi = \mu * N$ to obtain

$$\sum_{n\le x} \phi(n) = \sum_{n\le x} n \sum_{m|n} \frac{\mu(m)}{m} = \sum_{m\le x} \mu(m) \sum_{l\le x/m} l$$

and use a good approximation to the inner sum. □

Likewise the sum of two squares function

Theorem 0.13 (Gauss). *Suppose that $x \in \mathbb{R}$ and $x \ge 2$. Then*

$$\sum_{n\le X} r(n) = \pi X + O(X^{1/2}).$$

Again we leave the proof as an exercise. As a hint, one can again consider it as a lattice point problem, this time then number of lattice points inside a closed circle centre the origin and of radius $\sqrt{x}$. Then, there is a general principal which is easy to prove in this case that the number of lattice points in a convex region is equal to the area of the region with an error proportional to the length of the boundary. One way of seeing this is to associate the square $[u, u+1) \times [v, v+1)$ with the lattice point u, v and to observe that all the relevant lattice points are inside the circle radius $\sqrt{x}+\sqrt{2}$ and their union contains the circle radius $\sqrt{x}-\sqrt{2}$.

0.4. **Elementary Prime number theory.** The strongest results we know about the distribution of primes use complex analytic methods. However there are some very useful and basic results that can be established elementarily. Many expositions of the results we are going to describe use nothing more than properties of binomial coefficients, but it is good to start to get the flavour of more sophisticated methods even though here they could be interpreted as just properties of binomial coefficients. We start by introducing

The von Mangold function. This is defined by

$$\Lambda(n) = \begin{cases} 0 & \text{if } p_1p_2|n \text{ with } p_1 \neq p_2, \\ \log p & \text{if } n = p^k. \end{cases}$$

The interesting thing is that the support of Λ is on the prime powers, the higher powers are quite rare, at most $\sqrt{x}$ of them not exceeding x.

This function is definitely not multiplicative, since $\Lambda(1) = 0$. However it does satisfy some interesting relationships.

Lemma 0.14. *Let $n \in \mathbb{N}$. Then $\sum_{m|n} \Lambda(m) = \log n$.*

Proof. Write $n = p_1^{k_1} \dots p_r^{k_r}$ with the p_j distinct. Then for a non-zero contribution to the sum we have $m = p_s^{j_s}$ for some

s with $1 \le s \le r$ and j_s with $1 \le j_s \le k_s$. Thus the sum is

$$\sum_{s=1}^{r} \sum_{j_s=1}^{k_s} \log p_s = \log n.$$

□

We need to know something about the average of $\log n$.

Lemma 0.15 (Stirling)**.** *Suppose that $X \in \mathbb{R}$ and $X \ge 2$. Then*

$$\sum_{n \le X} \log n = X(\log X - 1) + O(\log X).$$

This can be thought of as the logarithm of Stirling's formula for $\lfloor X \rfloor!$.

Proof. We have

$$\begin{aligned}\sum_{n \le X} &= \sum_{n \le X} \left(\log X - \int_n^X \frac{dt}{t} \right) \\ &= \lfloor X \rfloor \log X - \int_1^X \frac{\lfloor t \rfloor}{t} dt \\ &= X(\log X - 1) + \int_1^X \frac{t - \lfloor t \rfloor}{t} dt + O(\log X).\end{aligned}$$

□

Now we can say something about averages of the von Mangoldt function.

Theorem 0.16. *Suppose that $X \in \mathbb{R}$ and $X \ge 2$. Then*

$$\sum_{m \le X} \Lambda(m) \left\lfloor \frac{X}{m} \right\rfloor = X(\log X - 1) + O(\log X).$$

Proof. We substitute from the first lemma into the second. Thus

$$\sum_{n \le X} \sum_{m|n} \Lambda(m) = X(\log X - 1) + O(\log X).$$

Now we interchange the order in the double sum and count the number of multiples of m not exceeding X. □

At this stage it is necessary to introduce some of the fundamental counting functions of prime number theory. For $X \geq 0$ we define

$$\psi(X) = \sum_{n \leq X} \Lambda(n),$$
$$\vartheta(X) = \sum_{p \leq X} \log p,$$
$$\pi(X) = \sum_{p \leq X} 1.$$

The following theorem shows the close relationship between these three functions.

Theorem 0.17. *Suppose that $X \geq 2$. Then*

$$\psi(X) = \sum_{k} \vartheta(X^{1/k}),$$
$$\vartheta(X) = \sum_{k} \mu(k) \psi(X^{1/k}),$$
$$\pi(X) = \frac{\vartheta(X)}{\log X} + \int_2^X \frac{\vartheta(t)}{t \log^2 t} dt,$$
$$\vartheta(X) = \pi(X) \log X - \int_2^X \frac{\pi(t)}{t} dt.$$

Note that each of these functions are 0 when $X < 2$, so the sums are all finite.

Proof. By the definition of Λ we have

$$\psi(X) = \sum_{k} \sum_{p \leq X^{1/k}} \log p = \sum_{k} \vartheta(X^{1/k}).$$

Hence we have

$$\sum_k \mu(k)\psi(X^{1/k}) = \sum_k \mu(k) \sum_l \vartheta(X^{1/(kl)}).$$

Collecting together the terms for which $kl = m$ for a given m this becomes

$$\sum_m \vartheta(X^{1/m}) \sum_{k|m} \mu(k) = \vartheta(X).$$

We also have

$$\begin{aligned}\pi(X) &= \sum_{p \le X} (\log p) \left(\frac{1}{\log X} + \int_p^X \frac{dt}{t \log^2 t} \right) \\ &= \frac{\vartheta(X)}{\log X} + \int_2^X \frac{\vartheta(t)}{t \log^2 t} dt.\end{aligned}$$

The final identity is similar.

$$\vartheta(X) = \sum_{p \le X} \log X - \sum_{p \le X} \int_p^X \frac{dt}{t}$$

etcetera. □

Now we come to a series of theorems which are still used frequently.

Theorem 0.18 (Chebyshev)**.** *There are positive constants C_1 and C_2 such that for each $X \in \mathbb{R}$ with $X \ge 2$ we have*

$$C_1 X < \psi(X) < C_2 X.$$

Proof. For any $\theta \in \mathbb{R}$ let

$$f(\theta) = \lfloor \theta \rfloor - 2 \left\lfloor \frac{\theta}{2} \right\rfloor.$$

Then f is periodic with period 2 and

$$f(\theta) = \begin{cases} 0 & (0 \le \theta < 1), \\ 1 & (1 \le \theta < 2). \end{cases}$$

Hence

$$\begin{aligned}\psi(X) &\geq \sum_{n\leq X}\Lambda(n)f(X/n)\\ &= \sum_{n\leq X}\Lambda(n)\left\lfloor\frac{X}{n}\right\rfloor - 2\sum_{n\leq X/2}\Lambda(n)\left\lfloor\frac{X/2}{n}\right\rfloor.\end{aligned}$$

Here we used the fact that there is no contribution to the second sum when $X/2 < n \leq X$. Now we apply Theorem 0.16 and obtain for $x \geq 4$

$$\begin{aligned}X(\log X-1) - 2\frac{X}{2}\left(\log\frac{X}{2}-1)\right) + O(\log X)&\\ = X\log 2 + O(\log X).&\end{aligned}$$

This establishes the first inequality of the theorem for all $X > C$ for some positive constant C. Since $\psi(X) \geq \log 2$ for all $X \geq 2$ the conclusion follows if C_1 is small enough.

We also have, for $X \geq 4$,

$$\psi(X) - \psi(X/2) \leq \sum_{n\leq X}\Lambda(n)f(X/n)$$

and we have already seen that this is

$$X\log 2 + O(\log X).$$

Hence for some positive constant C we have, for all $X > 0$,

$$\psi(X) - \psi(X/2) \leq CX.$$

Hence, for any $k \geq 0$,

$$\psi(X2^{-k}) - \psi(X2^{-k-1}) < CX2^{-k}.$$

Summing over all k gives the desired upper bound. □

The following now follow easily from the last couple of theorems.

Corollary 0.19 (Chebyshev)**.** *There are positive constants C_3, C_4, C_5, C_6 such that for every $X \ge 2$ we have*

$$\begin{aligned} C_3X <&\vartheta(X) < C_4X, \\ \frac{C_5X}{\log X} <&\pi(X) < \frac{C_6X}{\log X}. \end{aligned}$$

It is also possible to establish a more precise version of Euler's result on the primes.

Theorem 0.20 (Mertens)**.** *There is a constant B such that whenever $X \ge 2$ we have*

$$\begin{aligned} \sum_{n\le X} \frac{\Lambda(n)}{n} &= \log X + O(1), \\ \sum_{p\le X} \frac{\log p}{p} &= \log X + 0(1), \\ \sum_{p\le X} \frac{1}{p} &= \log\log X + B + O\left(\frac{1}{\log X}\right). \end{aligned}$$

Proof. By Theorem 0.16 we have

$$\sum_{m\le X} \Lambda(m) \left\lfloor \frac{X}{m} \right\rfloor = X(\log X - 1) + O(\log X).$$

The left hand side is

$$X \sum_{m\le X} \frac{\Lambda(m)}{m} + O(\psi(X)).$$

Hence by Cheyshev's theorem we have

$$X \sum_{m\le X} \frac{\Lambda(m)}{m} = X\log X + O(X).$$

Dividing by X gives the first result.

We also have

$$\sum_{m\le X}\frac{\Lambda(m)}{m}=\sum_k\sum_{p^k\le X}\frac{\log p}{p^k}.$$

The terms with $k\ge 2$ contribute

$$\le\sum_p\sum_{k\ge 2}\frac{\log p}{p^k}\le\sum_{n=2}^{\infty}\frac{\log n}{n(n-1)}$$

which is convergent, and this gives the second expression.

Finally we can see that

$$\begin{aligned}\sum_{p\le X}\frac{1}{p}&=\sum_{p\le X}\frac{\log p}{p}\left(\frac{1}{\log X}+\int_p^X\frac{dt}{t\log^2 t}\right)\\&=\frac{1}{\log X}\sum_{p\le X}\frac{\log p}{p}+\int_2^X\sum_{p\le t}\frac{\log p}{p}\frac{dt}{t\log^2 t}.\end{aligned}$$

Let

$$E(t)=\sum_{p\le t}\frac{\log p}{p}-\log t$$

so that by the second part of the theorem we have $E(t)\ll 1$. Then the above is

$$\begin{aligned}=&\frac{\log X+E(X)}{\log X}+\int_2^X\frac{\log t+E(t)}{t\log^2 t}dt\\=&\log\log X+1-\log\log 2+\int_2^{\infty}\frac{E(t)}{t\log^2 t}dt\\&+\frac{E(X)}{\log X}-\int_X^{\infty}\frac{E(t)}{t\log^2 t}dt.\end{aligned}$$

The first integral here converges and the last two terms are

$$\ll\frac{1}{\log X}.$$

□

There is an interesting application of the above which lead to some important developments. As a companion to the definition of a multiplicative function we have

Definition. An $f \in \mathcal{A}$ is **additive** when it satisfies $f(mn) = f(m) + f(n)$ whenever $(m, n) = 1$.

Now we introduce two further functions.

Definition. We define $\omega(n)$ to be the number of different prime factors of n and $\Omega(n)$ to be the total number of prime factors of n.

Example. We have $360 = 2^3 3^2 5$ so that $\omega(360) = 3$ and $\Omega(360) = 6$. Generally, when the p_j are distinct, $\omega(p_1^{k_1} \dots p_r^{k_r}) = r$ and $\Omega(p_1^{k_1} \dots p_r^{k_r}) = k_1 + \cdots k_r$.

One might expect that most of the time Ω is appreciably bigger than ω, but in fact this is not so. By the way, there is some connection with the divisor function. It is not hard to show that

$$2^{\omega(n)} \le d(n) \le 2^{\Omega(n)}.$$

In fact this is a simple consequence of the chain of inequalities

$$2 \le k + 1 \le 2^k.$$

Theorem 0.21. *Suppose that $X \ge 2$. Then*

$$\sum_{n \le X} \omega(n) = X \log\log X + BX + O\left(\frac{X}{\log X}\right)$$

where B is the constant of Theorem 0.20, and

$$\sum_{n\le X}\Omega(n) =$$

$$X\log X\log X + \left(B+\sum_p \frac{1}{p(p-1)}\right)X + O\left(\frac{X}{\log X}\right).$$

Proof. We have

$$\begin{aligned}\sum_{n\le X}\omega(n) &= \sum_{n\le X}\sum_{p|n}1 = \sum_{p\le X}\left\lfloor\frac{X}{p}\right\rfloor\\ &= X\sum_{p\le X}\frac{1}{p}+O\big(\pi(x)\big)\end{aligned}$$

and the result follows by combining Corollary 0.19 and Theorem 0.20.

The case of Ω is similar. We have

$$\sum_{n\le X}\Omega(n) = X\sum_{\substack{p,k\\ p^k\le X}}\frac{1}{p^k}+O\left(\sum_{k\le(\log X)/(\log 2)}\pi(X^{1/k})\right).$$

When $k\ge 2$ the terms in the error are $\ll X^{1/2}$ and so the total contribution from the $k\ge 2$ is $\ll X^{1/2}\log X$. In the main term, when $k\ge 2$ it remains to understand the behaviour of

$$\sum_{k\ge 2}\sum_{p>X^{1/k}}\frac{1}{p^k}\le\sum_{p>X^{1/2}}\frac{1}{p^2}+\sum_{k\ge 3}\frac{1}{(X^{1/k})^{k/2}}\sum_p\frac{1}{p^{k/2}}$$

The first sum is $\ll X^{-1/2}$ and the second is

$$\ll X^{-1/2}\sum_p\frac{1}{p(p^{1/2}-1)}\ll X^{-1/2}.$$

□

Hardy and Ramanujan made the remarkable discovery that $\log\log n$ is not just the average of $\omega(n)$, but is its normal order. Later Turán found a simple proof of this.

Theorem 0.22 (Hardy & Ramanujan)**.** *Suppose that $X \geq 2$. Then*

$$\sum_{n\le X}\left(\omega(n)-\sum_{p\le X}\frac{1}{p}\right)^2 \ll X\sum_{p\le X}\frac{1}{p},$$

$$\sum_{n\le X}(\omega(n)-\log\log X)^2 \ll X\log\log X$$

and

$$\sum_{2\le n\le X}(\omega(n)-\log\log n)^2 \ll X\log\log X$$

Turán. It is easily seen that

$$\sum_{n\le X}\left(\sum_{p\le X}\frac{1}{p}-\log\log X)\right)^2 \ll X$$

and (generally if $Y \geq 1$ we have $\log Y \leq 2Y^{1/2}$)

$$\begin{aligned}\sum_{2\le n\le X}(\log\log X-\log\log n)^2 &= \sum_{2\le n\le X}\left(\log\frac{\log X}{\log n}\right)^2\\ &\ll \sum_{n\le X}\frac{\log X}{\log n}\\ &= \sum_{n\le X}\int_n^X\frac{dt}{t}\\ &= \int_1^X\frac{\lfloor t\rfloor}{t}dt\\ &\le X.\end{aligned}$$

Thus it suffices to prove the second statement in the theorem. We have

$$\begin{aligned}\sum_{n\le X}\omega(n)^2 &= \sum_{p_1\le X}\sum_{\substack{p_2\le X\\ p_2\ne p_1}}\left\lfloor\frac{X}{p_1p_2}\right\rfloor+\sum_{p\le X}\left\lfloor\frac{X}{p}\right\rfloor\\ &\le X(\log\log X)^2+O(X\log\log X).\end{aligned}$$

Hence

$$\begin{aligned}\sum_{n\le X}(\omega(n)-\log\log X)^2 \le{}& 2X(\log\log X)^2\\ &-2(\log\log X)\sum_{n\le X}\omega(n)+O(X\log\log X)\end{aligned}$$

and this is $\ll X\log\log X$. □

One way of interpreting this theorem is to think of it probabilistically. It is saying that the events $p|n$ are approximately independent and occur with probability $\frac{1}{p}$. One might guess that the distribution is normal, and this indeed is true and was established by Erdős and Kac about 1941. Let

$$\Phi(a,b)=\lim_{x\to\infty}\frac{1}{x}\mathrm{card}\{n\le x : a<\frac{\omega(n)-\log\log n}{\sqrt{\log\log n}}\le b\}.$$

Then

$$\Phi(a,b)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}dt.$$

The proof uses sieve theory, which we might explore later.

0.5. **Orders of magnitude of arithmetical functions.** It is sometimes useful to know something about the way that an arithmetical function grows. Multiplicative functions tend to oscillate quite a bit in size. For example

$d(p) = 2$ but if we take n to be the product of the first k primes, say

$$n = \prod_{p \le X} p$$

for some large X, then

$$\log n = \vartheta(X)$$

so that

$$X \ll \log n \ll X$$

by Chebyshev and so

$$\log X \sim \log\log n,$$

but

$$d(n) = 2^{\pi(X)}$$

so that

$$\begin{aligned} \log d(n) &= (\log 2)\pi(X) \\ &\ge (\log 2)\frac{\vartheta(X)}{\log X} \\ &\sim (\log 2)\frac{\log n}{\log\log n}. \end{aligned}$$

Theorem 0.23. *For every $\varepsilon > 0$ there are infinitely many n such that*

$$d(n) > \exp\left(\frac{(\log 2 - \varepsilon)\log n}{\log\log n}\right).$$

The function $d(n)$ also arises in comparisons, for example in deciding the convergence of certain important series. Thus it is useful to have a simple universal upper bound.

Theorem 0.24. *Let $\varepsilon > 0$. Then there is a positive number C which depends at most on ε such that for every $n \in \mathbb{N}$ we have*

$$d(n) < Cn^{\varepsilon}.$$

Note, such a statement is often written as

$$d(n) = O_\varepsilon(n^\varepsilon)$$

or

$$d(n) \ll_\varepsilon n^\varepsilon.$$

Proof. Note that it suffices to prove the theorem when

$$\varepsilon \le \frac{1}{\log 2}.$$

Write $n = p_1^{k_1} \dots p_r^{k_r}$ where the p_j are distinct. Recall that

$$d(n) = (k_1 + 1) \dots (k_r + 1).$$

Thus

$$\frac{d(n)}{n^\varepsilon} = \prod_{j=1}^{r} \frac{k_j + 1}{p_j^{\varepsilon k_j}}.$$

Since we are only interested in an upper bound the terms for which $p_j^\varepsilon > 2$ can be thrown away since $2^k \ge k + 1$. However there are only $\le 2^{1/\varepsilon}$ primes p_j for which

$$p_j^\varepsilon \le 2.$$

Morever for any such prime we have

$$p_j^{\varepsilon k_j} \ge 2^{\varepsilon k_j} = \exp(\varepsilon k_j \log 2) \ge 1 + \varepsilon k_j \log 2 \ge (k_j + 1)\varepsilon \log 2.$$

Thus

$$\frac{d(n)}{n^\varepsilon} \le \left(\frac{1}{\varepsilon \log 2}\right)^{2^{1/\varepsilon}}.$$

□

The above proof can be refined so as to give a companion to Theorem 0.23

Theorem 0.25. *Let $\varepsilon > 0$. Then for all sufficiently large n we have*

$$d(n) < \exp\left(\frac{(\log 2 + \varepsilon)\log n}{\log\log n}\right).$$

Proof. We follow the proof of the previous theorem until the final inequality and then make then replace the ε there with

$$\frac{(1+\varepsilon/2)\log 2}{\log\log n}$$

which for large n certainly meets the requirement of being no larger than $1/\log 2$. Now

$$\begin{aligned}\left(\frac{1}{\varepsilon\log 2}\right)^{2^{1/\varepsilon}} &= \exp\left(\exp\left(\frac{\log\log n}{1+\varepsilon/2}\right)\log\frac{\log\log n}{(1+\varepsilon/2)\log 2}\right)\\ &< \exp\left(\frac{\varepsilon(\log n)\log 2}{2\log\log n}\right)\end{aligned}$$

for sufficiently large n. Hence

$$\begin{aligned}d(n) &< n^{\frac{(1+\varepsilon/2)\log 2}{\log\log n}}\exp\left(\frac{\varepsilon(\log n)\log 2}{2\log\log n}\right)\\ &= \exp\left(\frac{(1+\varepsilon)(\log n)\log 2}{\log\log n}\right)\\ &< \exp\left(\frac{(\log 2+\varepsilon)(\log n)}{\log\log n}\right).\end{aligned}$$

□